基于介电特征的苹果霉心病检测方法

李东博[1]，黄铝文[1,2,3*]，赵旭博[4]

（1. 西北农林科技大学 信息工程学院，陕西杨凌 712100；2. 农业农村部农业物联网重点实验室，陕西杨凌 712100；3. 陕西省农业信息感知与智能服务重点实验室，陕西杨凌 712100；4. 西北农林科技大学 食品科学与工程学院，陕西杨凌 712100）

摘　要：针对苹果霉心病无法有效根据外表进行识别，且传统检测方法具有设备复杂、成本高昂等问题，本研究通过采集苹果介电参数构建苹果霉心病检测模型，从而实现简单快速的苹果霉心病无损检测。基于LCR测量仪采集220个苹果的108项介电指标（9个频率下的12项介电指标）作为原始参数，使用数据标准化、主成分分析算法等对数据进行预处理，并利用BP神经网络、支持向量机、随机森林算法构建霉心病果检测模型。试验结果表明，基于随机森林算法构建的霉心病果检测模型性能最佳，在150个苹果构建的训练集和70个苹果构建的测试集中分类准确率分别达到96.66%和95.71%；基于采用BP神经网络构建的霉心病果检测模型效果次之，分类准确率分别可达到94.66%和94.29%；基于使用支持向量机构建的模型检测效果相对较差，分类准确率分别为93.33%和91.43%。试验结果表明，使用随机森林构建的模型可以更有效地识别霉心病果和好果。本研究可为苹果病虫害及苹果品质无损检测等提供参考。

关键词：苹果霉心病；介电特征；随机森林；BP神经网络；支持向量机

中图分类号：S126　　**文献标志码：**A　　**文章编号：**202102-SA035

引用格式：李东博，黄铝文，赵旭博. 基于介电特征的苹果霉心病检测方法[J]. 智慧农业(中英文)，2021, 3 (4): 66-76.

LI Dongbo, HUANG Lyuwen, ZHAO Xubo. Detection method of apple mould core based on dielectric characteristics[J]. Smart Agriculture, 2021, 3 (4): 66-76. (in Chinese with English abstract)

1　引　言

苹果霉心病是一种严重的苹果采后病害，感染霉心病后的苹果会由果实的心室往外扩展，逐渐导致果肉的腐烂[1]，但观察其外表却无法发现明显的病害特征。患有霉心病的苹果含有展青霉素和曲棒霉素等毒素[2]，具有影响人体生育、致癌、致突变等毒理作用，对人体健康产生较大威胁。霉心病发病时主要作用苹果心室，无法有效使用肉眼分辨苹果是否患有霉心病。因此便捷快速地检测苹果是否感染霉心病对于苹果品质检测、苹果的采后管理等具有重要的现实意义。

有研究针对苹果内部品质检测尝试使用不同的方法，主要包括近红外光谱[3,4]、计算机视觉[5,6]、核磁共振[7,8]、介电特征[9,10]等。McGlone等[11]使用石英卤钨灯照射苹果，获取其近红外光谱，并据此构建出苹果褐变在线检测系

收稿日期：2021-02-20　修订日期：2021-03-11

基金项目：国家自然科学基金面上项目（31671780）；宁夏自治区重点研发计划项目（2017BY067）

作者简介：李东博（1996—），男，硕士研究生，研究方向为苹果无损检测和数据降维。E-mail：1078105837@qq.com。

*通信作者：黄铝文（1976—），男，博士，副教授，研究方向为生物图像处理、机器人控制技术。电话：13709223117。E-mail：huanglvwen@nwafu.edu.cn。

统，最优模型决定系数可达90%。Shenderey等[12]使用近红外光谱技术，构建苹果霉心病检测模型，试验结果表明霉心病检测成功率达到90.1%，其中霉心病果的识别准确率为88.4%，好果的识别准确率为91.7%。李顺峰等[13]通过采集苹果的近红外漫射光谱，并使用多元散射校正、标准正太变量变换等不同的预处理方法进行处理，之后使用主成分分析算法进行数据降维，从而构建出相应的检测模型，结果表明训练集识别准确率达89.9%，测试集识别准确率达87.8%。王富春等[14]使用基于计算机视觉的方法，设计了苹果霉心病病变程度的检测方法，试验结果表明误检率为8.87%，可以较为有效地测量苹果霉心病变程度。Stroshine[15]使用低频磁共振设备，并安装高速传感器对苹果进行无损检测，判别准确率达到88%以上。张立彬等[16]通过采集苹果的介电特征，并对苹果阻抗、相对介电常数和损耗因数等内部品质与介电特征之间的关系进行了研究，发现可以通过介电特征达到判别苹果内部特征的目的。李芳等[17]使用LCR测量仪采集苹果在100 Hz～3.98 MHz间的介电参数，并通过稀疏主元分析、支持向量机、人工神经网络等建模方法构建对应的分类器，经过十折交叉验证分析，结果表明对于霉心病果的识别准确率可达94%。

相较于其他检测方法，使用介电方式具有检测灵敏性高，设备简单、成本低、易于实现的特点，但在苹果霉心病无损检测领域却较少有人使用介电方式，且已有文献着重于分析苹果霉心病对介电参数的影响，所构建的苹果霉心病检测模型识别准确率也有进一步提高的空间。本研究通过增加霉心病果样本数量，采用与已有研究不同的数据预处理及建模方法，提高霉心病果检测正确率，保证所建模型的准确性和有效性。本研究使用数据标准化、主成分分析算法等进行介电数据预处理，使用BP神经网络、支持向量机和随机森林算法构建苹果霉心病检测模型，以进一步提高苹果霉心病的检测准确率。

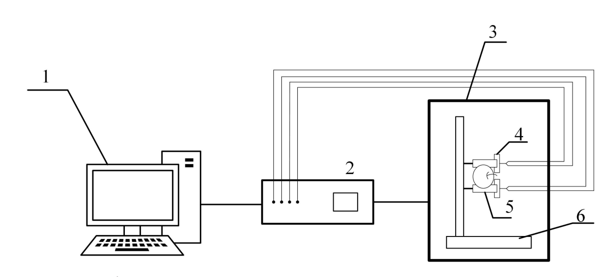
2 材料与方法

2.1 试验材料

本研究以苹果9种频率下的12项介电指标为研究对象，所使用苹果采集自陕西省白水县西北农林科技大学苹果试验站，品种为“富士”，通过人工精心选择挑选一批好果和疑似霉心病果。从中挑选大小相似、颜色相近、表皮无损伤的苹果220个，于室温下储藏。

2.2 苹果介电参数采集系统

图1所示为苹果介电参数采集系统，所使用主要设备为日本日置3532-50型LCR测试仪，测试探头型号为L2000 4端开尔文夹。



注：1. 计算机 2. 日本日置3532-50型LCR测试仪 3. 屏蔽箱 4. 平行电极板 5. 测试探头 6. 底座及支架

图1 苹果介电参数采集系统

Fig. 1 Acquisition system of apple dielectric parameter

为避免外界环境所导致的系统误差，采集苹果介电数据时首先将苹果置于室内24 h，待苹果温度与室内温度平衡后将苹果放入屏蔽箱，上下移动平行电极板使苹果紧贴于平行电极板之上，使用LCR测试仪探头夹住平行电极板进行测量，介电参数通过RS-232串行接口传输到电脑上。采集数据时沿果实最大横截面测量，以保持苹果与平行电极板间接触面积最大，保证所获取介电数据可以准确反映苹果内部品质。

苹果的理化性质与介电特性在频率为以10为底的对数下相关关系更加明显[18,19]，因此本研究使用158、251、398、15,800、25,100、39,800、1,580,000、2,510,000和3,980,000 Hz共9个频率，每个频率下采集12项介电指标，所采集介电指

标如表1所示。

表1 苹果12项介电参数说明

Table 1 Illustration of 12 dielectric parameters

介电参数编号	符号	解释	单位
1	*Z*	复阻抗	Ω
2	*Cs*	串联电容	F
3	*Cp*	并联电容	F
4	*Rs*	串联电阻	Ω
5	*Rp*	并联电阻	Ω
6	*Ls*	串联电感	H
7	*Lp*	并联电感	H
8	*G*	电导	S
9	*Y*	导纳	S
10	*θ*	复阻抗相角	(°)
11	*D*	损耗系数	——
12	*Q*	品质因子	——

采集数据时，分别为每个苹果进行编号，每项介电指标将其频率作为下标，如Z_{158}表示该苹果在频率为158 Hz下所采集的复阻抗，其他介电指标同理可得。同时为避免采集过程中系统性误差对试验效果产生影响，每个苹果分别采集5次数据，最后以5次数据的平均值作为该苹果的介电数据。

采集完成电学指标之后，对每个苹果进行破坏性试验，观察其是否为霉心病果，最终确认霉心病果164个，好果56个。划分训练集与测试集时，随机选取110个霉心病果和40个好果作为训练集，剩余的54个霉心病果和16个好果作为测试集进行后续试验。

2.3 原始数据预处理

使用LCR仪采集到介电参数，由于数据格式是以二维表的形式存储，即对于每一个苹果而言，其介电数据都可视为9 × 12的二维数组，对于所有数据可视为220 × 9 × 12的三维矩阵。但是此数据格式不利于进行数据处理，因此需要进行数据扁平化操作。与此同时，所采集的介电参数是在较大范围频率下采集的，数据分布范围不均匀，数据大小有明显差异。不同的介电指标具有不同的量纲，导致采集到的介电指标数据相差较大。为防止试验中大数据对小数据的影响，消除不同指标间的量纲影响，需要对原始数据进行标准化处理。

进行数据扁平化处理时，分别将每个苹果的108项指标由9 × 12的二维矩阵转换为1 × 108的一维矩阵，同时将所有的苹果数据进行拼接，最终形成220 × 108的苹果介电数据库。

苹果的介电参数是在9种不同频率下采集的，因此相同介电指标在不同频率下数据会产生较大的差异，如图2所示。在同一频率中不同的介电指标由于各自量纲的不同，数据的大小也会存在较大差异，如图3所示。

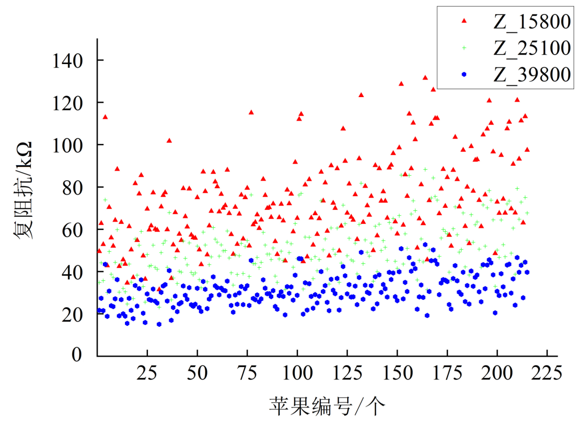
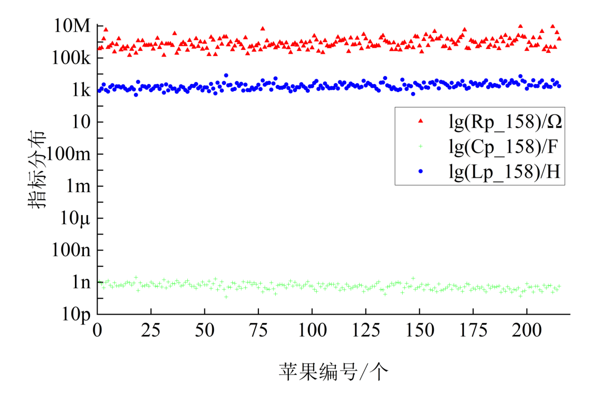


图2 不同频率下复阻抗分布趋势

Fig. 2 Distribution trend of complex impedance at different frequencies



注：纵坐标使用对数表示

图3 相同频率下不同介电指标分布趋势

Fig. 3 Distribution trend of different dielectric indices at the same frequency

图2中以复阻抗为例展示了其在15,800、25,100和39,800 Hz中的分布趋势，从中可以观察到不同频率对复阻抗具有较大影响，随着频率

的上升，复阻抗值以倍数形式扩增。图3中展示了158 Hz下并联电阻、并联电感及并联电容的变化情况，因为三者之间数据相差巨大，所以使用对数刻度的样式进行展示，可以观察到三者处于不同分布，相互之间具有数量级差异。

由于数据间的巨大差异，不进行数据预处理将有较大概率导致后续试验中较大的数据训练时占据更大的比重，较小的数据则占据更小的比重甚至于直接被忽略。而较小的数据中也可能蕴藏着较为重要的信息，因此需要将原始数据进行处理，消除不同量纲的影响。

预处理时使用了Z-score标准化方法对原始数据进行处理，其数学模型如下所示。

$$x^* = \frac{x - \bar{x}}{\sigma} \tag{1}$$

其中，x^*为标准化后的数据；x为原始数据；$\bar{x}$为原始数据均值；σ为原始数据标准差标准化后数据，如图4所示。

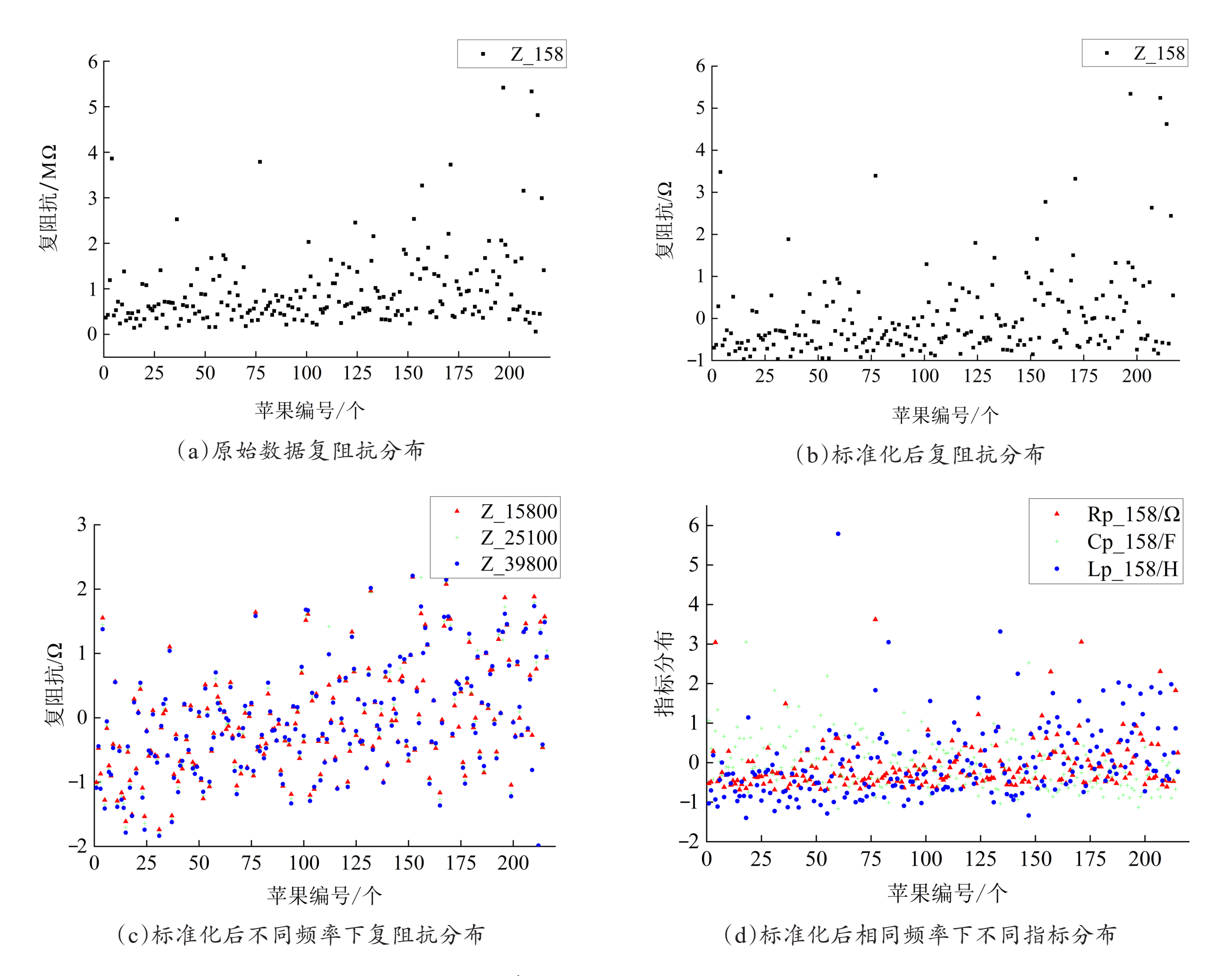


图4 介电参数标准化前后数据分布对比

Fig. 4 Distribution comparison of dielectric parameters before and after standardization

图4（a）是原始数据中Z_{158}的数据分布情况，图4（b）是经过标准化后Z_{158}的数据分布情况。两者对比可以看出两者数据分布几乎没有变化，但是数据分布区间已经大大减小，同时数据间的关联性得到了很好的保留，证明了标准化后的数据依旧有效。通过对比图2与图4（c）可以得到，标准化后的数据可以很好地将不同频率下的数据映射到较小的区间，同时可以保留数据间的关联性。同理，观察图3与图4（d）可得，经标准化后，数据可以消除不同指标间度量单位带来

的影响，并可很好地保留原始数据的原有信息。

2.4 建模方法

介电特性数据分析方法常用的算法有BP神经网络（Back Propagation Neural Network，BPNN）预测模型[20]、支持向量机（Support Vector Machine，SVM）[21]和随机森林（Random Forest，RF）[22]等模型。BPNN是一种按照数据间误差进行逆向传播训练网络的多层前馈神经网络，是应用最广泛的神经网络之一，具有优秀的非线性映射能力，在众多领域有着不俗的表现。SVM是一种按照监督学习方式对数据进行分类的模型，可以使用核方法进行非线性映射，在分类领域占据着重要的地位。RF是采取了集成思想的算法，将多个弱分类器通过组合变成一个强分类器。RF通过随机选取不同的训练样本和数据特征，根据每种训练样本和数据特征构建出对应的决策树，通过统计森林中所有决策树的分类结果，从而得出最终的分类结果，在多种数据集中都拥有着优秀的表现，且相较于其他算法可以有效降低训练时长。

本研究根据经预处理并使用主成分分析（Principal Component Analysis，PCA）[23]算法降维后所得主成分作为模型输入，结合BPNN、SVM以及RF构建苹果霉心病分类模型，对三种建模方法分别构建对应的分类模型，通过对比分析确定最优的模型。

3 结果与分析

3.1 介电数据降维

数据采集时通过9个不同的频率采集苹果的12项介电参数，因此对于每个苹果而言，不同频率下的某些介电指标经过标准化后蕴藏的信息量可能会极其相似，导致信息冗余，因此需要使用降维算法消除数据间的信息冗余。

将标准化之后的介电数据采用PCA方法降维后，其前30项主成分贡献率占比如图5所示。

从图中可以观察到，第十主成分之后提供的

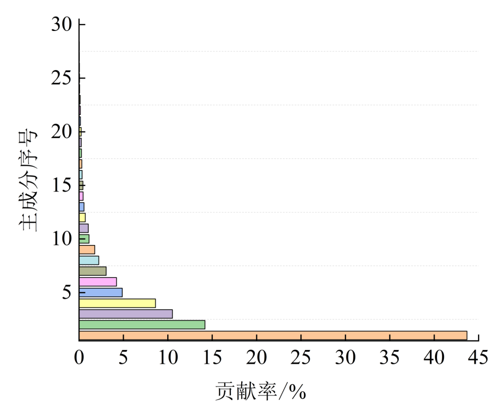


图5 PCA各主成分贡献率占比

Fig. 5 PCA contribution ratio of principal components

贡献率已经极其微小，结合具体数值分析，第一主成分（PC1）贡献率达到43.69%，第二主成分（PC2）、第三主成分（PC3）和第四主成分（PC4）的贡献率分别达到了14.19%、10.53%和8.61%，前四个主成分累计贡献率已经达到近80%。但是考虑到如果选取的主成分数量较少，可能会导致后续试验中因为训练数据过少而出现欠拟合现象，因此选取降维后的前14个主成分，其贡献率以及累计贡献率如表2所示。

表2 PCA降维后14个主成分贡献率及累计贡献率

Table 2 Contribution rate and cumulative contribution rates of 14 principal components after dimension reduction of PCA

主成分编号	PC1	PC2	PC3	PC4	PC5	PC6	PC7
贡献率/%	43.69	14.19	10.53	8.61	4.87	4.23	3.06
累计贡献率/%	43.69	57.88	68.41	77.02	81.89	86.12	89.18
主成分编号	PC8	PC9	PC10	PC11	PC12	PC13	PC14
贡献率/%	2.22	1.77	1.13	1.04	0.71	0.56	0.46
累计贡献率/%	91.40	93.17	94.31	95.34	96.05	96.61	97.07

从表2中可以观察到前五个主成分累计贡献率已经达到81.89%，前十个主成分累计贡献率达到94.31%，从第十四主成分（PC14）开始，后续主成分所提供的贡献率渐渐小于0.5%，已经较难提供更多原始数据中的信息。考虑主成分的个数以及累计贡献率之间的关联，为了后续试验分类的准确性，选取累计贡献率达到97.07%的主成分，即前14个主成分。

3.2 建模方法与结果

试验中采用BPNN、SVM与RF三种建模方法构建霉心病果、好果分类模型，三组试验中输入特征变量都为经PCA处理后的主成分变量，三组试验彼此独立进行，对比模型分类效果后选择分类效果最理想的模型作为最终分类模型。为后续表述简洁，三组试验分别以PCA-BP、PCA-SVM和PCA-RF表示。

3.2.1 PCA-BP模型参数的选择与设定

本研究所使用的BP神经网络为三层BP神经网络，包含输入层、隐藏层以及输出层。该模型中使用Sigmoid函数作为传递函数，影响该模型分类效果的因素包括隐藏层神经元个数、初始权值、学习速率以及期望误差。增加隐藏层神经元的个数相较于增加隐藏层的层数从结构实现角度而言更加简单，也可降低模型的复杂度。隐藏层神经元太少时，BP神经网络无法很好地进行学习，迭代次数也会延长，模型精度不够。隐藏层神经元个数太多时，网络的学习时间会逐步增加，同时也容易出现过拟合现象。初始权值的选择会影响后续网络训练的复杂度，合理的初始权值会有效降低模型训练时间，反之不合理的初始权值会增加模型训练代价。学习速率同样会对模型的训练及精度产生影响，过大的学习精度可能会导致模型不稳定，同时也有可能会跳过最优解，过小的学习精度会导致收敛速度过慢，从而需要花费更多的训练时间。

模型中如采用固定值作为初始权值，则会导致隐藏层节点都具有相同的计算结果，进而导致模型的收敛速度将大幅降低。因此为保证模型的收敛速度，防止出现梯度消失或者梯度爆炸现象，因此将初始权值设定为（-1，1）之间的随机数，期望误差设置为0.001，隐藏层节点个数按照经验公式，并结合试凑法确定最佳隐藏层节点数。经验公式如下所示。

$$h = \sqrt{m + n} + \alpha \tag{2}$$

其中，h为隐藏层节点数量，个；m为输入层节点数量，个；n为输出层节点数，个；α为1～10之间的常数。

试验中输入层节点个数设置为15，输出层节点个数设置为1，α取值设置为5，因此隐藏层节点个数设置为9。

考虑到模型中不同阶段会使用到不同的学习速率，因此设置学习速率时并没有固定学习速率，而是根据变学习率公式进行动态调节，变学习率公式如下所示：

$$\eta(t) = \eta_{max} - t \times (\eta_{max} - \eta_{min})/t_{max} \tag{3}$$

其中，η_{max}为最大学习率；η_{min}为最小学习率；t_{max}为最大迭代次数，次；t为当前迭代次数，次。

本试验中最大学习率设置为0.1，最小学习率设置为0.0001，最大迭代次数设置为3000。

3.2.2 PCA-SVM模型参数的选择与设定

SVM通过将原始数据进行映射，使其在高维空间中线性可分，影响SVM分类效果的参数主要有核函数、多项式维度、停止训练的误差值大小、最大迭代次数、是否采用概率估计、惩罚参数C、核函数Gamma参数等，其中对试验结果影响最大的有核函数类型、Gamma以及C。径向基核函数（Radical Basis Function，RBF）可以实现非线性映射，减少数据计算难度，同时RBF所涉及的参数较少，相较于其他多项式核函数可以有效降低模型的复杂度，因此本研究中选择RBF作为核函数。C与Gamma的选取采用了网格搜索化方法，该方法只需将参数输入进去，即可自行进行调优，选取最优的参数。本研究将C和Gamma的原始参数分别设置为［0.001，0.01，0.1，0.5，1，1.5，2，3，4，5，10，100，1000］和［0.001，0.005，0.0001，0.0005，0.00001，0.00005］，经过采用十折交叉验证法进行相关试验之后，最终得到的最优参数组合为C取3，Gamma取0.0005。

3.2.3 PCA-RF模型参数的选择与设定

随机森林是由一系列决策树构成的强分类器，其分类能力的强弱取决于所构建的决策树的分类能力，主要影响的参数包括节点分裂时所需

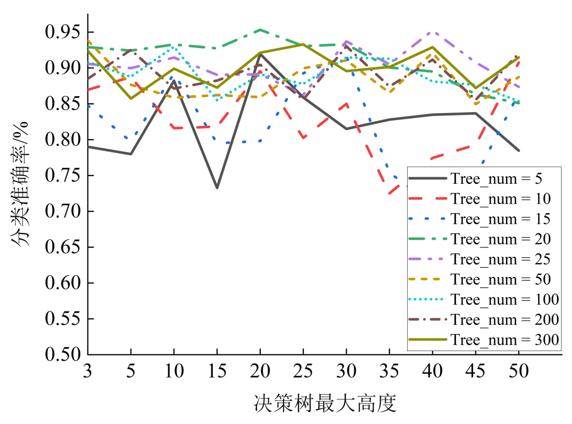
的最小样本数量（min_split），叶子节点最少样本数量（min_leaf），分裂所需最小增益（min_split_gain），采样方式（sampleing_mode），随机森林中决策树数量（tree_num），决策树高度（tree_depth）等。经预试验最佳参数组合如表3所示。

表3 RF部分参数选择

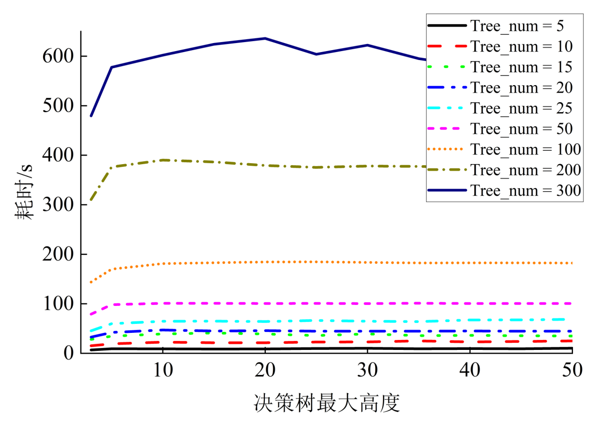
Table 3 Partial parameter selection of random forest

参数名称	min_split	min_leaf	min_split_gain	sampleing_mode
参数取值	6	2	$10e^{-6}$	sqrt

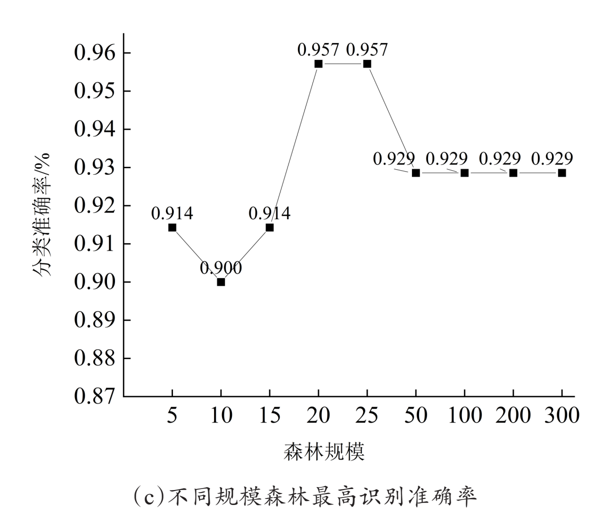
同时为了更加深入比较决策树数量及决策树高度对分类效果的影响，将森林规模分别设置为：5，10，15，20，25，50，100，200，300；每棵决策树的最大深度设置为：3，5，10，15，20，25，30，35，40，45，50。试验中在每种森林规模下分别将决策树设置为不同的深度，通过比较每种不同规模森林及不同决策树高度下分类的准确率及训练时间，从而选取最为合适的森林规模和决策树高度。不同规模森林及不同决策树高度分类准确率如图6所示。



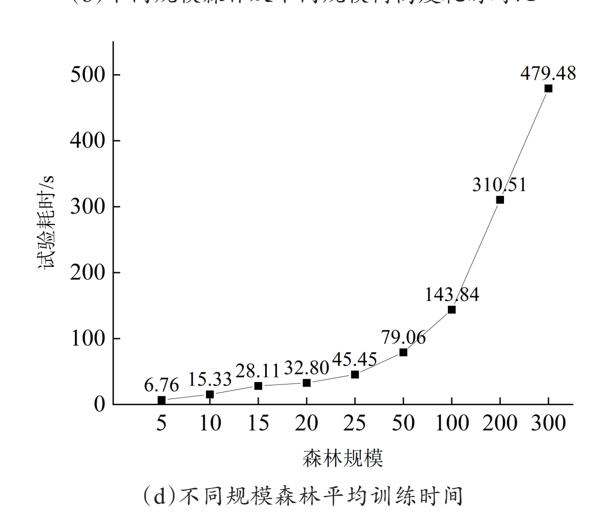
（a）不同规模森林及不同树高度分类准确率对比



（b）不同规模森林及不同规模树高度耗时对比



（c）不同规模森林最高识别准确率



（d）不同规模森林平均训练时间

图6 RF森林规模及树深度对分类效果的影响

Fig. 6 Effects of forest size and tree depth on the classification of random forest

从图6（a）中可以观察到，当森林规模控制在20和25时都可以获取较为不错的分类准确率，最高分类准确率都可达到95%以上。当森林规模控制到5、10和15时识别准确率较低，在不同决策树深度下，识别准确率有较大的波动。与此同时，当森林规模较高时，同样可以获取较为准确

的识别准确率。但是从图6（b）中可以看出，随着森林规模的增加，搭建分类模型所需的时间将逐步增加，当森林的规模超过50时，训练所需的时间将大幅上升。所以将森林的规模控制在50以内会是较好的选择。结合图6（a）与图6（b）可得到，虽然森林规模较高时分类模型同样具有较好的分类准确性，但是其训练耗时远远超过森林规模较小时的训练耗时。因此，为了避免漫长的训练耗时，应该选择较小的森林规模。

为了更加清晰地比较不同规模森林的识别准确率和耗时，提取了每种森林规模下的最好分类准确率和平均试验耗时，如图6（c）和图6（d）所示。从图中可以观察到，当森林规模为20和25时拥有着最高的识别准确率，同时训练耗时也较少。因此应该选择森林规模和决策树深度分别为（20，20）和（25，40）作为下一步试验参数，以期望获得更好的分类效果。

为进一步获取更加准确的试验结果，根据上文中取得的表现较好的森林规模和决策树深度，进行十折交叉验证试验。将原始数据分成相等的十份，每次取其中的一份作为测试集，将剩余九份作为训练集，最终取十次结果的平均值作为最终的分类准确率，试验结果如图7所示。

从图7可以明显看出，当森林规模为20、决策树深度为20时，分类模型的分类准确率要优

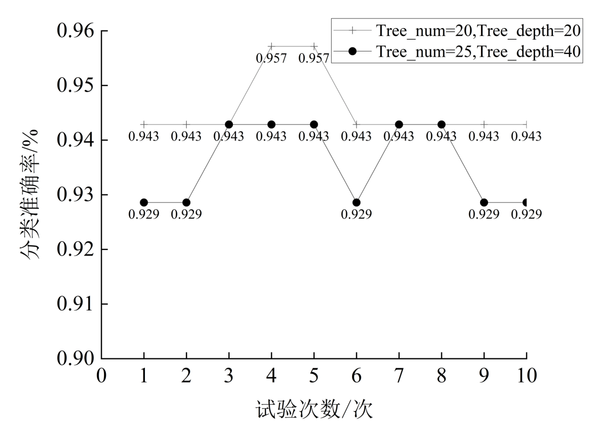


图7 随机森林十折交叉验证分类准确率

Fig.7 Ten-fold cross validation classification accuracy of random forest

于另一组，两者的平均识别准确率分别为0.946和0.936。因此，在森林规模为20、决策树深度为20时可以取得最优的分类效果，最高分类准确率可达0.957。

3.2.4 模型结果分析

将原始苹果介电参数经过PCA降维得到的主成分与经过破坏性试验所得到的苹果标签（其中标签为“1”表示为霉心病果，标签为“0”表示为正常果）组合起来作为模型输入变量，分别输入PCA-BP，PCA-SVM和PCA-RF三种模型中，三组模型的最优识别准确率结果如表4所示。

表4 三种模型识别准确率对比

Table 4 Recognition accuracy of three models

建模方法	训练集误判数		分类准确率/%	测试集误判数		分类准确率/%
	好果（40个）	霉心果（110个）		好果（16个）	霉心果（54个）	
PCA-SVM	2	8	93.33	2	4	91.43
PCA-BP	2	4	94.66	1	3	94.29
PCA-RF	1	4	96.66	0	3	95.71

从试验结果上可以观察到，PCA-RF的分类效果最好，训练集中识别准确率达到96.66%，训练集中150个苹果的错分样本数只有5个，误判率仅为3.33%。PCA-RF在共70个苹果组成的测试集中也有着最好的表现，分类准确率达到95.71%。错分样本共计3个，其中，好果全部可以正确分类，霉心病果中有3个样本被错分为好果。PCA-BP模型的分类效果仅次于PCA-RF，其在训练集中分类准确率达到94.66%，测试集中分类准确率为94.29%，训练集中错分样本为6

个，测试集中错分样本为4个，分类效果与PCA-BP相似。PCA-SVM模型分类效果相对较差，其在训练集中分类准确率可达93.33%，但在测试集中分类准确率仅能达到91.43%，错分样本量在训练集和测试集中分别为10个和6个，误判率分别为6.66%和8.57%。

从表4可以观察到三组模型出现误判时将霉心病果误判为好果的比例要大于将好果误判为霉心病果的比例，三组数据组合起来，出现误判的样本共计34条，其中霉心病果误判为好果占据26条，占比达到76.5%，而将好果误判为霉心病果的样本数共计8条，所占比例为23.5%。具体到每个模型中，以PCA-RF为例，训练集和测试集中错分样本共计8条，其中霉心病果错分为好果的数量为7条，其比例达到87.5%。总体而言，将霉心病果错分为好果的概率要远大于将好果错分为霉心病果的概率，导致这种现象出现的原因可能是由于样本中存在一些霉心果的症状较为轻微，还未大幅影响苹果的介电特征，所以导致模型分类时将其错分为好果。

4 结 论

本研究针对霉心病果分类这一具体问题，使用BPNN、SVM和RF三种建模方法，结合数据标准化、主成分分析算法等预处理方法构建出利用苹果9种频率下12项介电指标进行训练的霉心病果分类模型。试验结果表明，使用RF构建的模型分类效果更佳，在测试集中最高分类效果达到96.66%，BPNN所构建模型效果次之，测试集中分类效果达到94.29%，SVM所构建模型分类效果相对其余两种方法效果较差，测试集中分类效果仅为91.43%。

本研究构建的苹果霉心病分类模型可以有效地分辨出霉心病果和好果，有效地将苹果介电参数与苹果霉心病检测结合到一起，可为后续研究奠定基础。

与已有研究相比，本研究中使用了不同的数据预处理及建模方法，所构建的苹果霉心病检测模型更为有效，霉心病果的识别准确率有了较为明显的提高，但与此同时本研究仅针对“富士”这一种苹果，未对比其他苹果品种的分类准确率，在接下来的研究中可以采集不同品种苹果介电参数，对比不同品种下模型效果，提高模型通用性。

参考文献：

[1] PATRIARCA A. Fungi and mycotoxin problems in the apple industry[J]. Current Opinion in Food Science, 2019, 29: 42-47.

[2] ZHONG L, CARERE J, LU Z, et al. Patulin in apples and apple-based food products: The burdens and the mitigation strategies[J]. Toxins, 2018, 10(11): ID 475.

[3] 郭志明, 黄文倩, 陈全胜, 等. 近红外光谱的苹果内部品质在线检测模型优化[J]. 现代食品科技, 2016, 32(9): 147-153.
GUO Z, HUANG W, CHEN Q, et al. Optimization of online detection model for internal quality of apple by near infrared spectroscopy[J]. Modern Food Science and Technology, 2016, 32(9): 147-153.

[4] 王转卫, 迟茜, 郭文川, 等. 基于近红外光谱技术的发育后期苹果内部品质检测[J]. 农业机械学报, 2018, 49(5): 348-354.
WANG Z, CHI Q, GUO W, et al. Internal quality detection of apple in late development stage based on near infrared spectroscopy[J]. Transactions of the CSAM, 2018, 49(5): 348-354.

[5] 高辉, 马国峰, 刘伟杰. 基于机器视觉的苹果缺陷快速检测方法研究[J]. 食品与机械, 2020, 36(10): 125-129.
GAO H, MA G, LI W. Research on rapid detection method of apple defects based on machine vision[J]. Food & Machinery, 2020, 36(10): 125-129.

[6] BHARGAVA A, BANSAL A. Quality evaluation of Mono & bi-Colored apples with computer vision and multispectral imaging[J]. Multimedia Tools and Applications, 2020, 79(11): 7857-7874.

[7] 张京平, 朱建锡, 孙腾. 苹果内部品质的CT成像结合傅里叶变换方法检测[J]. 农业机械学报, 2014, 45(5): 197-204.
ZHANG J, ZHU J, SUN T. Detection of apple internal quality by CT imaging combined with Fourier transform method[J]. Transactions of the CSAM, 2014, 45

(5): 197-204.

[8] 曹玉栋, 祁伟彦, 李娴, 等. 苹果无损检测和品质分级技术研究进展及展望[J]. 智慧农业, 2019, 1(3): 29-45.
CAO Y, QI W, LI X, et al. Research progress and prospect on non-destructive detection and quality grading technology of apple[J]. Smart Agriculture, 2019, 1(3): 29-45.

[9] 蔡骋, 李永超, 马惠玲, 等. 基于介电特征选择的苹果内部品质无损分级[J]. 农业工程学报, 2013, 29(21): 279-287.
CAI C, LI Y, MA H, et al. Nondestructive grading of apple internal quality based on dielectric feature selection[J]. Transactions of the CSAE, 2013, 29(21): 279-287.

[10] 王若琳, 王栋, 任小林, 等. 基于电学特征的苹果水心病无损检测[J]. 农业工程学报, 2018, 34(5): 129-136.
WANG R, WANG D, REN X, et al. Nondestructive detection of apple water core disease based on electrical characteristics[J]. Transactions of the CSAE, 2018, 34(5): 129-136.

[11] MCGLONE V A, MARTINSEN P J, CLARK C J, et al. On-line detection of Brown heart in Braeburn apples using near infrared transmission measurements[J]. Postharvest Biology and Technology, 2005, 37(2): 56-63.

[12] SHENDEREY C, SHMULEVICH I, ALCHANATIS V, et al. NIRS detection of moldy core in apples[J]. Food & Bioprocess Technology, 2010, 3(1): 79-83.

[13] 李顺峰, 张丽华, 刘兴华, 等. 基于主成分分析的苹果霉心病近红外漫反射光谱判别[J]. 农业机械学报, 2011, 42(10): 158-161.
LI S, ZHANG L, LIU X, et al. Discrimination of apple moldy heart disease by near infrared diffuse reflectance spectroscopy based on principal component analysis[J]. Transactions of the CSAM, 2011, 42(10): 158-161.

[14] 王富春, 李军, 张润浩, 等. 基于计算机视觉的苹果霉心病病变程度测量方法[J]. 农机化研究, 2015, 37(6): 189-193.
WANG F, LI J, ZHANG R, et al. Measurement method of apple mold heart disease lesion degree based on computer vision[J]. Journal of Agricultural Mechanization Research, 2015, 37(6): 189-193.

[15] STROSHINE C R. Rapid sensing of internal browning in whole apples using a low-cost, low-field proton magnetic resonance sensor[J]. Postharvest Biology and Technology, 2005, 3: 56-64.

[16] 张立彬, 胥芳贾, 灿纯, 等. 苹果内部品质的电特性无损检测研究[J]. 农业工程学报, 2000(3): 104-106.
ZHANG L, XU F, CAN C, et al. Study on nondestructive testing of internal quality of Apple[J]. Transactions of the CSAE, 2000(3): 104-106.

[17] 李芳, 蔡骋, 马惠玲, 等. 基于生物阻抗特性分析的苹果霉心病无损检测[J]. 食品科学, 2013, 34(18): 197-202.
LI F, CAI C, MA H, et al. Nondestructive detection of apple moldy heart disease based on bioelectrical impedance analysis[J]. Food Science, 2013, 34(18): 197-202.

[18] 蔡骋, 李晓龙, 马惠玲, 等. 基于生物阻抗特性的苹果新鲜度无损测定[J]. 农业机械学报, 2013, 44(2): 147-152.
CAI C, LI X, MA H, et al. Nondestructive determination of apple freshness based on bioelectrical impedance characteristics[J]. Transactions of the CSAM, 2013, 44(2): 147-152.

[19] 耿晗, 蔡骋, 刘斌. 基于介电特性的苹果内部品质预测方法研究[J]. 农机化研究, 2019, 41(10): 186-191.
GENG H, CAI C, LIU B. Study on prediction method of apple internal quality based on dielectric properties[J]. Journal of Agricultural Mechanization Research, 2019, 41(10): 186-191.

[20] SIDHU K, GILL A, ARORA A, et al. Advancements in farming and related activities with the help of artificial intelligence: A review[J]. Environment Conservation Journal, 2021, 22(SE): 55-62.

[21] OMRANI E, KHOSHNEVISAN B, SABOOHI H, et al. Potential of radial basis function-based support vector regression for apple disease detection[J]. Measurement, 2014, 55: 512-519.

[22] SAMAJPATI B, DEGADWALA S. Hybrid approach for apple fruit diseases detection and classification using random forest classifier[C]// 2016 International Conference on Communication and Signal Processing (ICCSP). Piscataway, New York, USA: IEEE, 2016: 1015-1019.

[23] GRANATO D, SANTOS J, ESCHER G, et al. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties

in foods: A critical perspective[J]. Trends in Food Science & Technology, 2018, 72: 83-90.

Detection Method of Apple Mould Core Based on Dielectric Characteristics

LI Dongbo[1], HUANG Lyuwen[1,2,3*], ZHAO Xubo[4]

（1. *College of Information Engineering, Northwest A&F University*, *Yangling* 712100, *China*; 2. *Key Laboratory of Agricultural Internet of Things*, *Ministry of Agriculture and Rural Affairs*, *Yangling* 712100, *China*; 3. *Key Laboratory of Agricultural Information Perception and Intelligent Service*, *Yangling* 712100, *China*; 4. *College of Food Science and Engineering*, *Northwest A&F University*, *Yangling* 712100, *China*）

Abstract: Apple mouldy core disease often occurs in the ventricle of apples and cannot be effectively identified by appearance. Near-infrared spectroscopy, nuclear magnetic resonance and other methods are usually used in traditional apple mouldy core disease detection, but these methods require complex equipment and high detection costs. In this research, a simple and fast nondestructive detection method of apple mouldy core disease was proposed by using a dielectric method to construct an apple mouldy core disease detection model. Japan's Hioki 3532-50 LCR tester was used to collect 108 dielectric indicators (12 dielectric indicators at 9 frequencies) of 220 apples as the original data. Due to the large differences in the distribution of data collected with different dielectric indexes and different frequencies, a standardized method was used for data preprocessing to eliminate the problem of large differences in dielectric data distribution. Afterwards, in order to eliminate the redundant information between the data, the principal component analysis algorithm was used to reduce the data dimensionality, and finally the three algorithms of BP neural network (BPNN), support vector machine (SVM) and random forest (RF) were used to construct the mouldy core disease detection model. After pre-experiment, the most effective parameters of each algorithm were selected, the test results showed that the apple mouldy core disease detection model based on the RF algorithm obtained the best performance, and the detection accuracy rate reached 96.66% and 95.71% in the training set (150 apples) and the test set (70 apples). The mouldy core disease detection model constructed by using BPNN was the second most effective, and the detection accuracy could reach 94.66% and 94.29%, respectively. The detection effect of the model built by using SVM was relatively poor, and the detection accuracies were 93.33% and 91.43%, respectively. The experimental results showed that the model constructed by using RF can more effectively identify mouldy core disease apples and healthy apples. This study could provide references for apple diseases and insect pests and non-destructive testing of apple quality.

Key words: apple mouldy core disease; dielectric characteristics; random forest; BP neural network; support vector machine

（登陆 www.smartag.net.cn 免费获取电子版全文）